Auditory Spectral Integration Effects in Consonant-Vowel [bæ]-[dæ] Transitions

A Senior Honors Thesis

Presented in Partial Fulfillment of the Requirements for graduation with distinction in Speech and Hearing Sciences in the undergraduate colleges of The Ohio State University

by

Erin K. Saylor

The Ohio State University
June 2007

Project Advisors: Dr. Robert A. Fox and Dr. Ewa Jacewicz, Department of Speech and Hearing Science

Abstract

In speech, perceptual distinctions among consonants are typically made on the basis of formant transitions, or frequency changes in vocal resonances as voicing moves from consonant to vowel production. Previous studies have shown that these distinctions can be made whether frequency transitions are “real” or “virtual.” A real formant transition is an actual change or movement in frequency to or from the steady state vowel portion. A virtual transition refers to a change to the spectral Center-of-Gravity (COG), which can also be used to simulate formant transitions in consonants. Fox et al. (2007) demonstrated excellent identification along a /da-ga/ continuum using these simulated formant transitions. However, one important unanswered question is the level at which these formant transitions are processed. Are they processed peripherally by the cochlea or are they mediated more centrally by higher auditory structures? Two projects underway at The Ohio State University address this question. One project (Wackler, 2007) uses the same /da-ga/ continuum but presents the stimuli dichotically-- the consonant transition is presented in the left ear and the remainder of the stimulus is presented in the right ear, requiring central mediation in the identification task. The present study utilized a /bæ-dæ/ continuum, also presented in a dichotic paradigm. The difference in consonant continuum is important because the relevant information for distinguishing /bæ/ from /dæ/ is located at the F2 transition, as compared to the F3 transition for /da-ga/; in /bæ-dæ/, additional formants such as F3 do not carry information necessary for sound discrimination. The stimuli in the present study were constructed using only F1 and F2 components.

Twelve normal hearing adults (three males and nine females) between the ages of 20-27 participated in the study. They were asked to identify speech sounds in the /bæ-dæ/ continuum under four conditions: diotic synthetic speech, dichotic synthesized speech, diotic virtual speech and dichotic virtual speech signals. Preliminary studies, Wackler (2007) and the present [bæ]-[dæ] research, have shown parallel results demonstrated by Fox et al. (2007). A substantial number of subjects had a more difficult time discriminating virtual speech tokens in the dichotic condition while performance for synthetic speech tokens was similar for diotic and dichotic presentation. These results suggest that the COG formant transition may be as not salient in the second formant position, the presence of F3 information may provide additional spectral richness beneficial for discrimination and lastly, parameters of the stimuli such as intensity need to be further analyzed. Future studies will address these possibilities in order to clarify the nature of peripheral vs. central processing of virtual formant transitions. Results have implications for the development of improved cochlear implants for hearing-impaired persons.

Acknowledgements

I would like to thank and acknowledge Dr. Robert Fox and Dr. Ewa Jacewicz for allowing me the opportunity to be apart of their research study as well as their guidance and support. I would like to thank Chiung-Yun Chang for all of the time and effort she has given me through out the entire process. In addition I would also like to thank my co-laborer, Lisa Wackler for all of her help and encouragement. Finally I would like to acknowledge and sincerely thank Dr. Janet Weisenberger for always opening her door for any questions and concerns I had regarding my research and life in general. Her patience, understanding and willingness to always help as made this a truly valuable experience. I would like to extend my gratitude and appreciation to all of my subjects who volunteered their time, making it possible to conduct my research. Lastly, I would like to thank my family for their understanding and constant support.

The present study was supported by an ASC Undergraduate Research Scholarship and by the SBC Undergraduate Research Scholarship.

Table of Contents

Chapter 1: Introduction and Literature Review

Listening for most individuals is an effortless activity that requires little concentration. A fundamental aspect of listening is how we translate what we hear into meaningful information, that is, how we perceive sound. Speech perception is the process in which sounds observed in communication are interpreted. It involves the auditory system's ability to extract, integrate and decode auditory information embedded inside sound waves.

The speech production process involves sound generation via various coordinated muscle movements within the larynx that collectively act upon air flowing through the respiratory passages. The glottis, tongue, lips, jaw and velum operate simultaneously to produce vibratory sound patterns by modifying the primary airflow. The characteristics of the resultant sounds vary due to differences in individual vocal tract configuration, which directly affect a sound's resonance (Pickett, 1999). In vowel perception, the identification of speech sounds is based on concentrations of acoustic energy in particular frequency regions, known as formants, representing the resonances of the vocal tract. Speech determinants, such as consonants, can be identified on the basis of the first two formants F1 transitions into and out of steady-state vowel formants. Vowels, however, are identified primarily on the basis of the first two formants, F1 and F2. In speech perception, energy at these formants translates into activity on the basilar membrane in the cochlea, giving rise to the perception of specific speech sounds (Pickett, 1999).

Analyzing the elements of complex sound waves has provided essential information about acoustic characteristics such as formants. In doing so it has also

surfaced a greater need for understanding how the auditory system makes use of these unique features for distinguishing different speech sounds used in understanding speech. While phoneticians are able to describe the process and results of speech production, much is still unknown about mechanisms involved in speech perception.

For research purposes, previous investigations have found it useful to employ synthetic speech stimuli to examine the workings of the auditory system in regards to integrating and perceiving speech-like tokens. These stimuli are constructed by simulating the acoustic resonances present in natural speech. Synthetic sounds are desirable for studies that wish to limit variability otherwise found in natural speech. In addition, synthetic speech allows precise manipulation of individual frequency, amplitude and timing characteristics that co-vary in natural speech. Current research aims to gain understanding of how the auditory system discriminates and detects thresholds as well as unique acoustic attributes as a way of distinguishing various complex sounds.

Early research speech perception focused mainly on distinctive acoustic features salient for determining vowel quality in one and two-formant vowels. Auditory spectral integration has been the term most commonly used in studying how the auditory system integrates or merges components of sound during speech perception tasks. Delattre et al. (1952) investigated the importance of formants in speech perception by observing the performance during one and two-formant vowel matching tasks by implementing synthetic tokens while modifying amplitude, frequency and intensity. The study showed that the phonetic quality of the simplified back vowels containing only the first two formants could be matched to a vowel consisting of only one formant. However, the

success of the match was contingent upon the specific relationship between the frequencies and amplitudes of the two close formants and the peak frequency of the single formant. Similar results were produced when the relative intensity ratio between the two close formants in these synthetic back vowels were modified, causing a change in the frequency of the single formant to which it was best matched.

These results suggested that the auditory system must average formant frequencies that fall relatively close in proximity in order to produce this effect. This occurrence is known today as the center-of-gravity effect (COG), which suggests that the auditory system performs an additional filtering of vowels beyond the cochlea. It is important to note that Delattre employed two sets of vowels; one set in which the F1 and F2 were relatively close in frequency (back vowels) and a second in which these two formants were significantly separated (front vowels). The study found that the latter did not yield as many successful matches. These results indicate that the COG did not occur for those formants with a large degree of separation, rather two frequencies were perceived instead of one centrally averaged frequency. Delattre's findings demonstrated the significance of formant location, specifically the distance separating formants. Early experiments exploring the COG effect addressed the question of the limits of formant separation that could still yield formant averaging. The results of the earlier COG effects led to several studies investigating the integration of formants during vowel identification in vowel matching experiments. Numerous studies were conducted by Chistovich and colleagues (e.g., Bedrov et al., 1978; Chistovich and Lublinskaja, 1979; Chistovich et al., 1979; Chistovich, 1985), which illustrated that predictable shift of the single formant, occurred when the two close formants fell within a

"critical distance" or "critical formant separation" of about 3.5 bark. It was suggested that within this critical distance, changes to the relative amplitude ratios between the two formants changed their combined spectral COG. During vowel matching tasks it was this spectral COG to which the frequency of the single formant was matched. These studies proposed that when two vowel formant peaks are separated by less than 3.5 bark[1], they are perceptually integrated into a single perceived peak ("perceptual formant"), (Chistovich *et al.,* 1979). The 3.5-bark critical distance indicates a possible limit on spectral integration in that the COG effect disappears with larger formant separation.

Early psychoacoustic investigations by Feth (1974) and Feth and O'Malley (1977) examined the aforementioned type of auditory filtering for non-speech stimuli. The research explored the spectral pitch of complex tones using two-component complex tone pairs that had identical envelopes but differed in fine structure (Voelcker, 1996a; b). The results agreed with the COG hypothesis of Chistovich *et al.*, showing a decrease in discriminability as the separation of the two components increased to 3.5 bark. The two-tone resolution results of Feth and O'Malley (1977), complemented the critical distance data from Delattre (1952) and Chistovich and Lublinskaja (1979) during vowel matching tasks. This analysis reinforces the idea of a possible common mechanism, i.e., an auditory spectral resolving power. Recently, Xu *et al.* (2004) confirmed both the results of Chistovich and Lublinskaja (1979) and Feth and O'Malley (1977) with a set of American English listeners who responded to the same two types of signals, two-formant synthetic vowels and complex two-tones. In their study, Xu *et al.*

[1] This value, referred to as one bark, is an empirically derived common value that expresses the relation between pitch and frequency (Ladefoged 2001, p.167). It is thought to represent the width of one of the presumed bank of adjacent filters used to model basilar membrane mechanics (Scharf, 1972)

concluded that both the complex-tone discriminability and the spectral integration limits reflected the same auditory spectral resolving power, and therefore further suggests that the auditory processing of complex auditory signals at the intermediate stage is the same for speech and non-speech signals.

One important limitation of the studies was that the signals implemented in the perceptual research thus far have been spectrally static, that is, the parameters of the signal remained constant for the entire duration of the sound. Actual speech signals are typically dynamic, with formant frequencies and amplitudes changing over time. It became clear that in order to obtain a more accurate account of integration effects, the perceptual attributes of dynamic signals, being more prevalent in speech, needed to be studied

In order to satisfy the need for examining integration effects in dynamic speech signals, Lublinskaja (1996) utilized two Russian diphthongal vowels to observe the auditory system's ability to attend to a dynamic spectral COG. The ratios of the amplitudes of two relatively closely spaced formants were modified over time in order to generate the dynamic perception of a non-stationary, diphthongal vowel. Lublinskaja's results showed that these virtual formant changes were successful in producing a diphthongal-vowel percept when the critical distance between the modified formants, F2 and F3, was longer for the diphthongal transitions than for static vowels, reaching about 4.2 bark. However, when the distance between F2 and F3 exceeded 4.2 bark the percept was that of a stationary vowel. Lublinskaja's research helped reveal information regarding integration effects during dynamic vowel perception. Findings from her study

evoked the question as to whether or not the same results could be repeated using dynamic consonant-vowel transitions.

Addressing this question, Feth *et al.* (2006) extended the investigations of the COG effect observed in diphthongal vowels to consonant-vowel (CV) transitions. The study sought to determine whether or not virtual frequency (VF), and frequency modulated (FM) transitions could be processed in the same manner, making them perceptually equivalent to a synthetic formant transition. The experiment looked specifically at the CV transitions in /da/ and /ga/. The results revealed no significant difference in the identification responses between stimulus types, revealing that the dynamic change caused by amplitude modulation is a phenomenon comparable to a frequency change. Feth *et al.* concluded that this "virtual" frequency change is processed similarly in acoustic and speech signals and suggest that processing occurs at more central level in the auditory system.

Another study by Fox *et al.* (2007) sought to examine the extent to which the perception of synthetic [da]-[ga] and [t^ha]-[k^ha] syllables could be cued by virtual formant transitions or virtual bursts brought about by spectral COG effects. Their data illustrated that listeners could accurately identify and distinguish consonants along the /da/-/ga/ continuum using these stimulated, virtual formant transitions. In addition, these dynamic cues, useful for indicating place of articulation, were found to be perceptually equivalent despite the variance in signal type. Given, now that signal type, virtual or actual, demonstrated parallel perceptual results, the question arises whether or not other acoustic characteristics are salient for speech signal distinctions.

A number of studies were (e.g. Blumstein and Stevens, 1979; 1980; Stevens and Blumstein, 1978; Furui, 1986; Liberman et al., Nittrouer, 1992; Ohde et al., 1995; Kewley-Port,1983) have concluded that there are several cues which contribute to the quality of a stop consonant percept; specifically, the release bursts and formant transition. Extensive research regarding the acoustic cues have suggest that either one alone is sufficient for successful stop perception. However, Lieberman and Blumstein (1988) note that is has not been made clear that the auditory system independently extracts these cues during speech perception, and it can also be suggested that the signal attributes may actually be perceived as a single integrated cue (Stevens and Blumstein (1978)).

As noted above Fox *et al.* (2007) looked at both virtual burst and virtual formant transitions conducted in the syllables [da]-[ga] and [t^ha]-[k^ha] . These formant transitions differ from the tokens used by Lublinskaja (1996) in that they occur over a shorter amount of time and at a faster rate. Fox et al. selected a syllabic context to better understand the effects of a dynamic COG because it underlines the importance of acoustic transitions that appear in coarticulatory effects in speech. The focus of this study, in contrast to previous work, was not the limits of spectral integration, but rather how the auditory system utilizes rapid changes in COG effects across a wide range of frequencies over time in order to make phonetic distinctions. The results showed that listeners were able to perceive “virtual F3 transitions” with similar success to that of actual F3 transitions. These data indicated that the two types of dynamic cues provide comparable phonetic information. However, one important aspect of the perceptual system that has yet to be addressed is the level at which auditory processing occurs: at

the auditory periphery (within the cochlea), or at higher auditory levels, which are centrally located. Determining the location of this spectral integration could have clinical application in improving the design of cochlear implant processors.

The present study addresses this question by employing a dichotic listening paradigm and utilizing a /bæ-dæ/ continuum to examine the level of spectral integration. The contrast in presentation condition between diotic (both ears hearing identical auditory stimuli simultaneously) and dichotic (presentation of two different auditory stimuli, simultaneously, one to each ear) was used to determine whether successful perceptual distinctions in both conditions could be obtained, which would support the idea of auditory processing at a central level. In contrast to the consonant continuum used by Fox et al. (2006), a [bæ]-[dæ] continuum was implemented in order to examine the extent of integration at lower frequencies (F1 and F2 formants only). The stimuli were chosen because relevant information for distinguishing speech sounds in this continuum is located at the F2 transition and additional formants such as F3 are excess energy that does not carry information necessary for sound discrimination. Thus, the present study further examines the level and role of spectral integration in order to obtain a more comprehensive understanding of auditory function during speech perception.

Chapter 2: Method

I. Stimuli

The synthetic sounds were constructed so that there were four different [bæ]-[dæ] series that varied in terms of two conditions; listening condition in which tokens were presented (diotic or dichotic) and the type of F2 transition within each token (actual or virtual F2 transition).

II. Diotic, Actual F2 Transition Series

The actual F2 transitions were generated by using a parallel version of the Klatt synthesizer, using the .kld option in HLsyn (Sensimetrics, 1997), with a sampling rate of 11025 Hz. Each token had a total duration of 250-ms, consisting of a 50-ms consonant transition portion and a 200-msec steady-state vowel portion. During the first 50 ms, F1 increased from 480 to 620 Hz and remained constant at 620 Hz during the final 200 ms steady-state vowel portion. The transition portion of F1 and the steady-state portion of F1 and F2 were identical in all nine-frequency steps. In addition the steady-state frequency of F2, being 1660 Hz, also remained unchanged and was implemented in each of the nine continua. However, the 50 ms transition portion of F2 was varied in terms of their F2 onset frequencies, which were also equally spaced. Nine different F2 onset frequencies were constructed, ranging 1300 Hz to 1852 Hz, with an offset frequency of 1660 Hz. The onset and offset frequencies of the nine F2 transitions are shown in Table 1.1. In addition, Figure 1.1 is a schematic representation of the diotic, actual F2 transition series.

Table 2.1 Onset and Offset frequencies of Actual F2 Transitions

Series Step	Transition Onset	Transition Offset
1	1300	1660
2	1369	1660
3	1438	1660
4	1507	1660
5	1576	1660
6	1645	1660
7	1714	1660
8	1783	1660
9	1852	1660

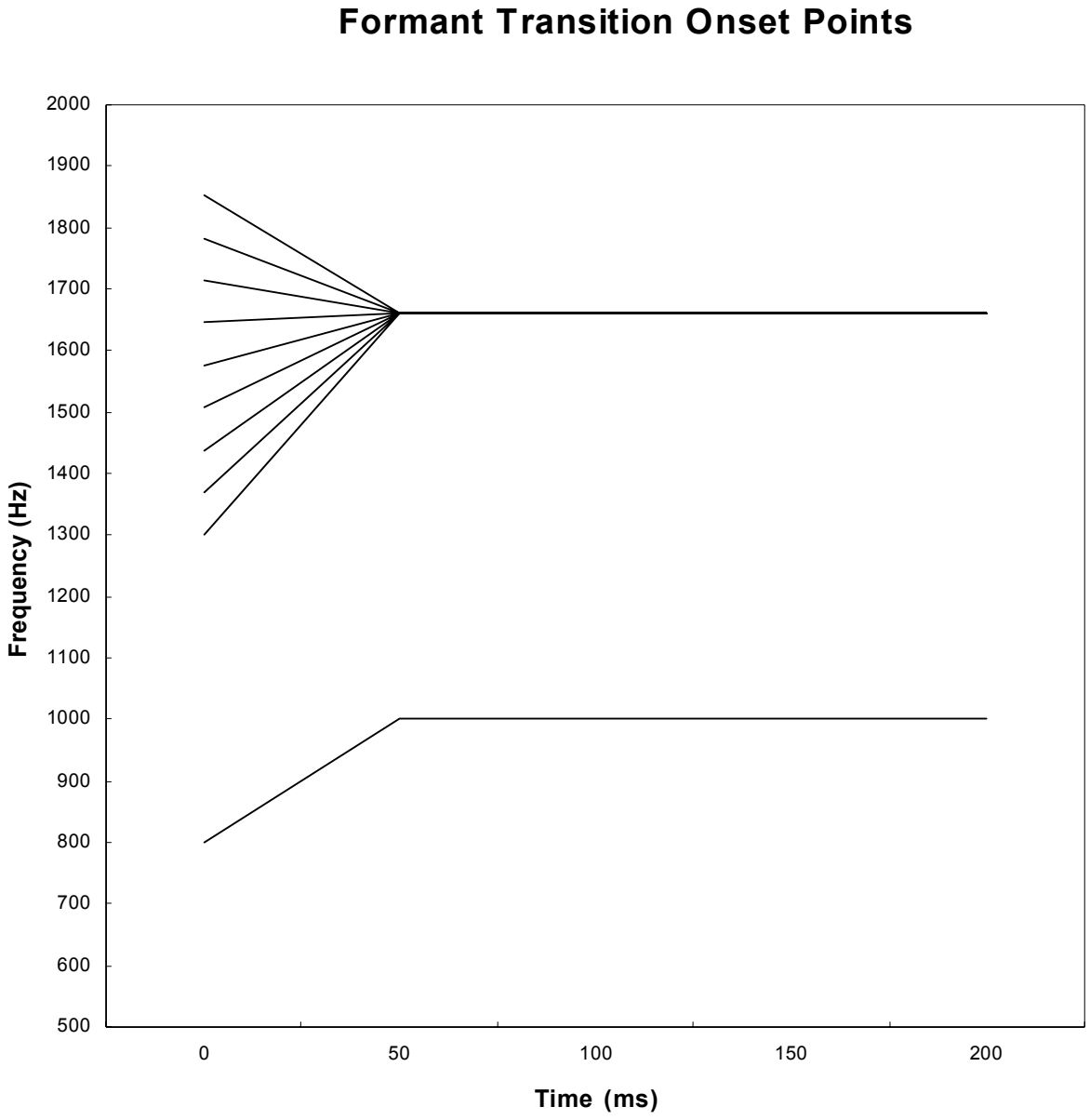

Figure 2.1 Schematic Representation of the diotic, actual F2 transition series; note that the frequency of the F2 onset varies from 1300 Hz to 1852 Hz, as outlined in Table 2.1

III. Diotic, Virtual F2 Transition Series

Using a two-step construction process, these stimuli were constructed as a virtual F2 transition that was inserted into the first 50 ms of consonant base-vowel token. The base token was created using the same parallel version of the klatt synthesizer, again using the HLsyn with .kld option and a sampling rate of 11025 Hz. The CV base was comprised of a 50 ms F1 transition as well as a 200 ms steady-state vowel portion for both F1 and F2, which were again identical in for each of the nine frequency transition steps. Over the first 50 msec, F1 increased from 480 Hz to 620 Hz and remained unchanged over the final 200 msec steady-state vowel portion of the token. The frequency of F2 steady state portion was a constant 1660 Hz. The nine different stimuli varied in terms of their 50-ms consonant transition. These transitions were generated by manipulating the relative intensity ratios of two sets of sinusoidal waveforms present at 1200 Hz and 1320 Hz (lower pair) and 1800 Hz and 1920 Hz (upper pair). The sine waves' frequencies were multiples of the fundamental frequency (F_o = 120 Hz). The sine wave pairs were separated by 2.6 bark in order to generate a center-of-gravity that fell just above or below the endpoints of the [bæ]-[dæ] frequency range. The sine wave pairs were created using the tone generator option in an Adobe Audition program and were constructed so that they were of equal amplitudes. The relative intensities of these sine wave pairs were systematically manipulated and combined in order to produce a virtual tonal glide. The average rms value of the integrated sine waves was adjusted to be within 0.1dB of the actual F2 transition. The intensity weighting used to create the virtual onsets and offsets are illustrated in Table 2.2

Table 2.2 Relative intensities of onsets and offsets of sine wave pairs. (lower pair harmonics used =1200 Hz and1320 Hz, mean =1260 Hz; upper pair, harmonics used = 1800 Hz and 1920 Hz, mean = 1860 Hz)

Series Step	Relative Intensity Onset (Lower Pair) 1200 Hz 1320 Hz	Relative Intensity Offset (Lower Pair) 1200 Hz 1320 Hz	Relative Intensity Onset (Upper Pair) 1800 Hz 1920 Hz	Relative Intensity Offset (Upper Pair) 1800 Hz 1920 Hz
1	93.33%	57.74%	6.67%	81.65%
2	81.83%	57.74%	18.17%	81.65%
3	70.33%	57.74%	29.67%	81.65%
4	58.83%	57.74%	47.17%	81.65%
5	47.33%	57.74%	52.67%	81.65%
6	35.83%	57.74%	64.17%	81.65%
7	24.33%	57.74%	75.67%	81.65%
8	12.83%	57.74%	87.17%	81.65%
9	1.33%	57.74%	98.67%	81.65%

Finally, after adjustments, the 50 ms virtual F2 transitions were inserted into a 250-ms time frame with the onset of the sine waves synchronized with the onsets of the F1 consonant-vowel base and F2 steady state portion via Adobe Audition. Varying the relative intensities values shown in Table 2.2 simulated a dynamic tonal glide, similar to that of an actual frequency transition. Increasing or decreasing the intensity of the upper and lower sine wave pairs over time resulted in a rising F2, generating the percept [dæ] or a falling F2, creating the percept [bæ]. A schematic of the diotic virtual F2 transitions is shown in Figure 2.2.

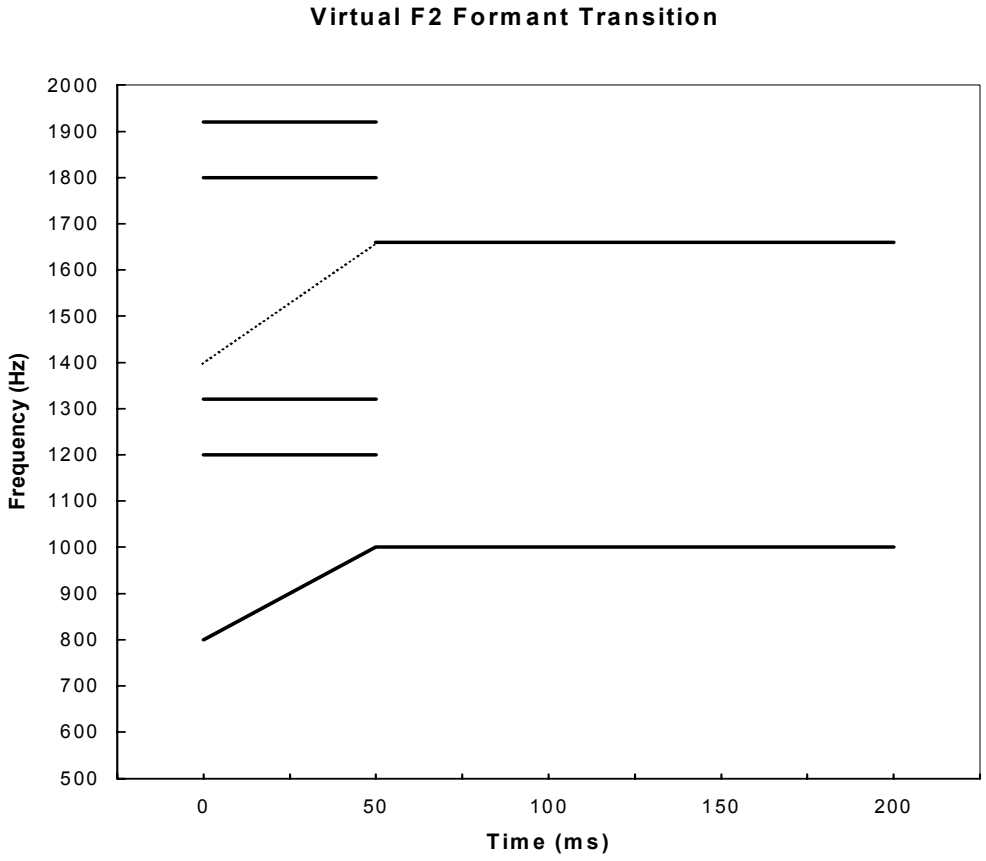

Figure 2.2 Schematic representation of the diotic, virtual F2 transition series; note that the center frequencies of the two 50-msec sine wave pairs remain constant, with the variation in frequency. The modified relative intensities, as outlined in Table 1.2, create the percept of a virtual F2 transition.

IV. Dichotic, Actual F2 Transitions

For the construction of this transition series Audition was used again, this time to implement a dichotic condition, meaning two separate channels were necessary for stimuli presentation. This allowed a selected portion of the token to be presented in one ear, while the remaining portion to be presented in the other ear. The same diotic, actual, F1 consonant-vowel base and F2 steady-state vowel portion with varying actual transition onsets were implemented in the dichotic condition as well. However, the CV base was inserted into the top channel while the F2 steady state with transitions was placed in the bottom channel. Over the first 50 msec, F1 increased from 480 Hz to 620 Hz and remained constant at 620 Hz. The F2 consonant transitions, created in the same manner as those previously described for the diotic condition, behaved in the same fashion, that is rising or falling in frequency to generate [bæ]-[dæ] percepts. A schematic of the dichotic actual F2 transitions are represented in Figure 2.3.

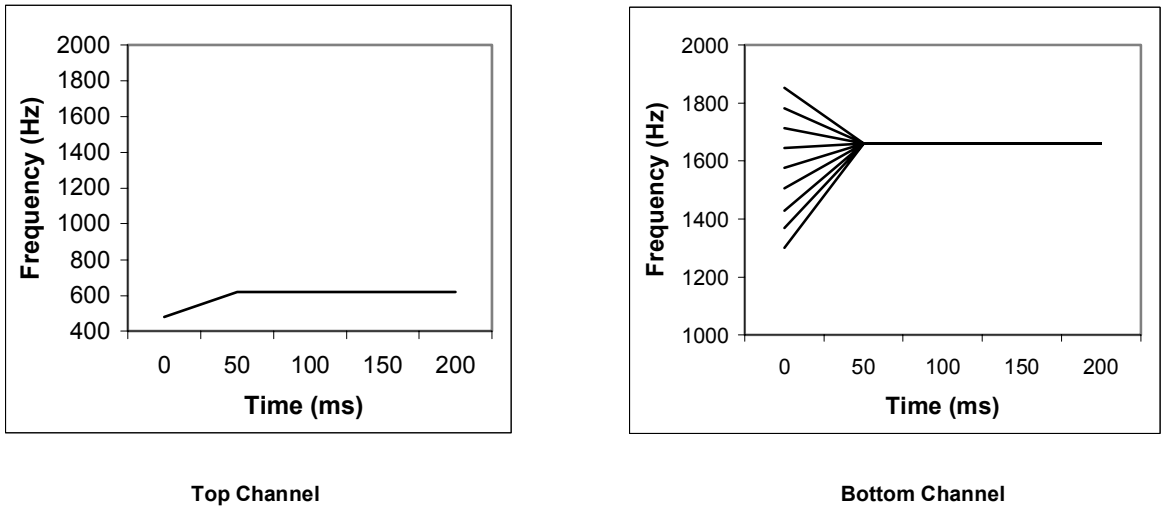

Figure 2.3. Schematic representation of the dichotic, actual F2 transition series; note that the F1 consonant-vowel base token is in the top channel and the F2 frequency onsets and steady-state portion is in the bottom channel.

IV. Dichotic, Virtual F2 Transitions

The construct for the dichotic virtual stimuli were similar to that of the dichotic, actual transitions, that is, part of the stimuli was inserted into one channel while the remaining portion of the stimuli was placed in the other channel. The F1 consonant-vowel base token was inserted into the top channel while the two sets of sine wave pairs and F2 steady-state vowel portion were placed in the bottom channel. Over the first 50 ms, F1 increased from 480 Hz to 620 Hz and remained constant at 620 Hz for the remaining 200 ms. The F2 portion of each stimulus consisted of a 50-ms virtual transition, created in the same manner as those used in the diotic, virtual F2 condition. That is, the relative intensities of the two sets sine wave pairs were manipulated in order to change the spectral center-of-gravity so that virtual transitions would produce a percept comparable to that of an actual F2 transition. Again, the final 200-ms of the F2 remained steady at 1660 Hz. Figure 2.4 is a schematic representation of the virtual F2 transition series in a dichotic condition.

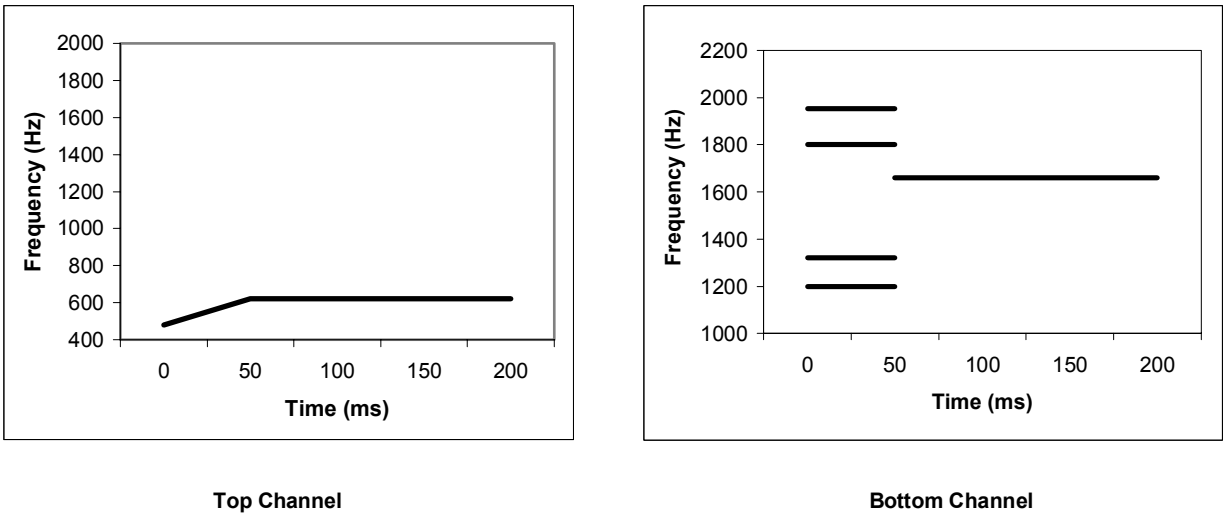

Figure 2.4. Schematic representation of the dichotic, virtual F2 transition series; note that the F1 consonant-vowel base token is in the top channel and the two pairs of harmonics and the F2 steady-state vowel portion is in the bottom channel.

V. Listeners

Twelve normal hearing adults (three males and nine females) between the ages of 20-25 participated in the study. All were students at The Ohio State University and native speakers of American English. The subjects were paid $16 for their participation.

VI. Procedure

Subjects listened to the stimuli in both diotic and dichtic presentation through Sennheiser HD600 headphones. The amplitude was set at a comfortable listening level (70 dB HL) unless otherwise requested by the participant. A single-interval 2AFC identification task was used which included the responses *bæ* and *dæ*. The two responses were displayed visually on two different halves of a divided computer monitor. The subjects were seated inside a sound-attenuating booth and were asked to indicate whether they heard *bæ* or *dæ* for each stimulus by clicking on the mouse button on the side of the screen that corresponded to that sound. There were 135 tokens randomly presented in each stimulus set (9 tokens x 15 repetitions), which were blocked by token type (diotic or dichotic tokens). For all listeners, diotic tokens were presented first, followed by dichotic tokens and within each block actual F2 transitions were played first, succeeded by virtual F2 transitions.

In order to familiarize listeners with the synthetic speech sounds and task procedure and decrease learning effects, the subjects prior to listening to each blocked stimulus continuum were presented with a training set consisting 25 tokens. The items were repetitions of the first and tenth step (the endpoints) of that specific series. During each preliminary set the stimulus was presented via headphones; shortly after the appropriate response would light up on the computer screen. The subjects listened to

the training examples twice and to the entire nine-step continuum once. Following the auditory training, subjects were asked to take a 15-item practice, with no feedback. After the practice was completed, the experimenter answered any questions or concerns the participant had before the actual test of 135 tokens. The experimental conditions included diotic and dichotic tokens, each consisting of two continua, “actual” and “virtual”. Each set lasted approximately 30 minutes. Two subjects were not able to complete the all four stimuli sets and were dismissed from the remainder of the study; their data were not included in the final analysis.

Chapter 3: Results and Discussion

Experiment 1: Sensitivity to COG Effects in the Diotic Condition

The first experiment aimed to determine listener's sensitivity to virtual F2 transitions generated by COG effects. Shown in Figure 3.1 are the identification function responses to both actual and virtual F2 transition. It is evident that the two identification functions have similar characteristics. The primary difference between the two functions is the gradation in slope. The actual formant transition series illustrates a relatively steep shift from mostly *bæ* responses to mostly *dæ* responses. However, the slope for the virtual F2 transition series is more gradual, generating a less distinctive shift in *bæ* to *dæ* responses.

An analysis of the 50% cross-over points (representing the location of the /b/-/d/ category boundary along the F2 onset axis for each individual) was obtained by using Probit analysis. The results revealed that there was no significant difference between the two functions ($F(1,9)=1.69$, $p>.05$, $\eta^2 =.16$); analysis of the total number of /d/ responses also revealed no significant difference ($F(1,11)=1.67$, $p>.05$, $\eta^2 =.13$). This data indicates that listeners demonstrated no unexpected response bias between the two series. Next, a two-way ANOVA with the within-subject factors F2 onset frequency (series step) and stimulus type (actual or virtual) of the number of /d/ responses was calculated. The analysis revealed a significant main effect of F2 onset frequency step, ($F(8,88)= 58.37$ $p<.001$, $\eta^2 =.84$), as expected but no significant main effect of stimulus type, ($F(1,11)= 1.67$, $p=.22$, $\eta^2 =.13$). The calculation also determined an interaction effect of F2 onset and stimulus type ($F(8,88)=10.23$, $p< .001$, $\eta^2=.48$). This interaction

stems from the fact that the slope of the ID function for the actual is significantly higher than the slope for the virtual formant transition ($t(11)=4.36$, $p=.001$). The results of experiment 1 can be interpreted that listeners can spectrally integrate the two sets of sine wave pairs to perceive a "virtual" formant transition that can cue the /b/-/d/ place distinction. The data also indicates that listeners can perceive the virtual F2 transitions as being parallel in providing similar phonetic information as actual F2 transition. However, due to the gradual slope of the virtual F2 ID function, it is implied that the virtual cue, even though effective in cueing /b/-/d/, is not as salient as the actual formant cue. The question that now arises is the level at which listeners spectrally integrate information in the auditory system.

Experiment 2: Examining Spectral Integration in the Dichotic Condition

The results of the first experiment support the claim that listeners are sensitive to virtual F2 transitions generated by COG effects and that listener can spectrally integrate actual and virtual F2 transitions with comparable success. In experiment 2, two further questions were addressed: Are listeners able to spectrally integrate actual F2 formant transitions dichotically and is the success of the COG effect in cueing /b/-/d/ contingent upon signal condition? Shown in Figure 3.2 are the identification function responses from both actual and virtual F2 transition series in the dichotic condition. Again the primary difference between the two functions is the disparity in slope. The location of the /b/-/d/ category boundary was found to have a cross over location relatively similar to that of actual F2 transition ID function in the diotic condition. The congruence leads

us to believe that the listening condition (diotic or dichotic) and the stimulus type (actual or virtual) do not greatly influence the listener's perceptual change in /b/-/d/. A two-way ANOVA with the within-subjects factors condition and F2 onset frequency step of the number of /d/ responses obtained showed that there was no significant main effect of condition ($F(1,7)= .19$, $p>.05$, $\eta^2=.03$), thus supporting the previous claim. Next, another two-way ANOVA with the within-subjects factors F2 onset frequency (series step) and stimulus type (actual or virtual) of the number of /d/ responses was calculated and no significant main effect of stimulus type ($F(1,8)= .048$, $p>.05$, $\eta^2 =.006$) was illustrated. However, as expected, a significant main effect of F2 onset frequency step, ($F(8,64)= 14.1$, $p<.001$, $\eta^2 =.64$) was shown. This calculation also established a significant interaction effect of F2 onset and stimulus type ($F(8,64)=7.12$, $p< .001$, $\eta^2=.47$).

An analysis of the 50% cross-over-points (representing the /b/-/d/ category boundary obtained from Probit analysis) indicated no significant difference between the two functions ($F(1,7)= .465$, $p>.05$, $\eta^2 =.06$); analysis of the total number of /d/ responses also revealed no significant difference ($F(1,8)= .001$, $p>.05$, $\eta^2 =.00$) Again this data indicates that listeners demonstrated no unexpected response bias between the two series. The difference in slopes between the ID functions in the dichotic conditions were even more striking than in the diotic condition with, again, the slope of the ID function for the actual transition being significantly higher than for the virtual condition ($t(8),=4.283$, $p=.003$).

Again we interpret these results to mean that listeners can spectrally integrate the two sets of sine wave pairs to perceive a "virtual" formant transition that can cue the

/b/-/d/ place distinction although the virtual cue is not as salient as the actual formant cue. In addition, t-tests indicated that although the slopes of the actual formant ID functions were not significantly different in the diotic vs. dichotic conditions ($t(8)=2.00$, $p>.05$), the slopes were significantly higher for the virtual condition in the diotic vs. the dichotic conditions ($t(8)=4.52$, $p<.002$).

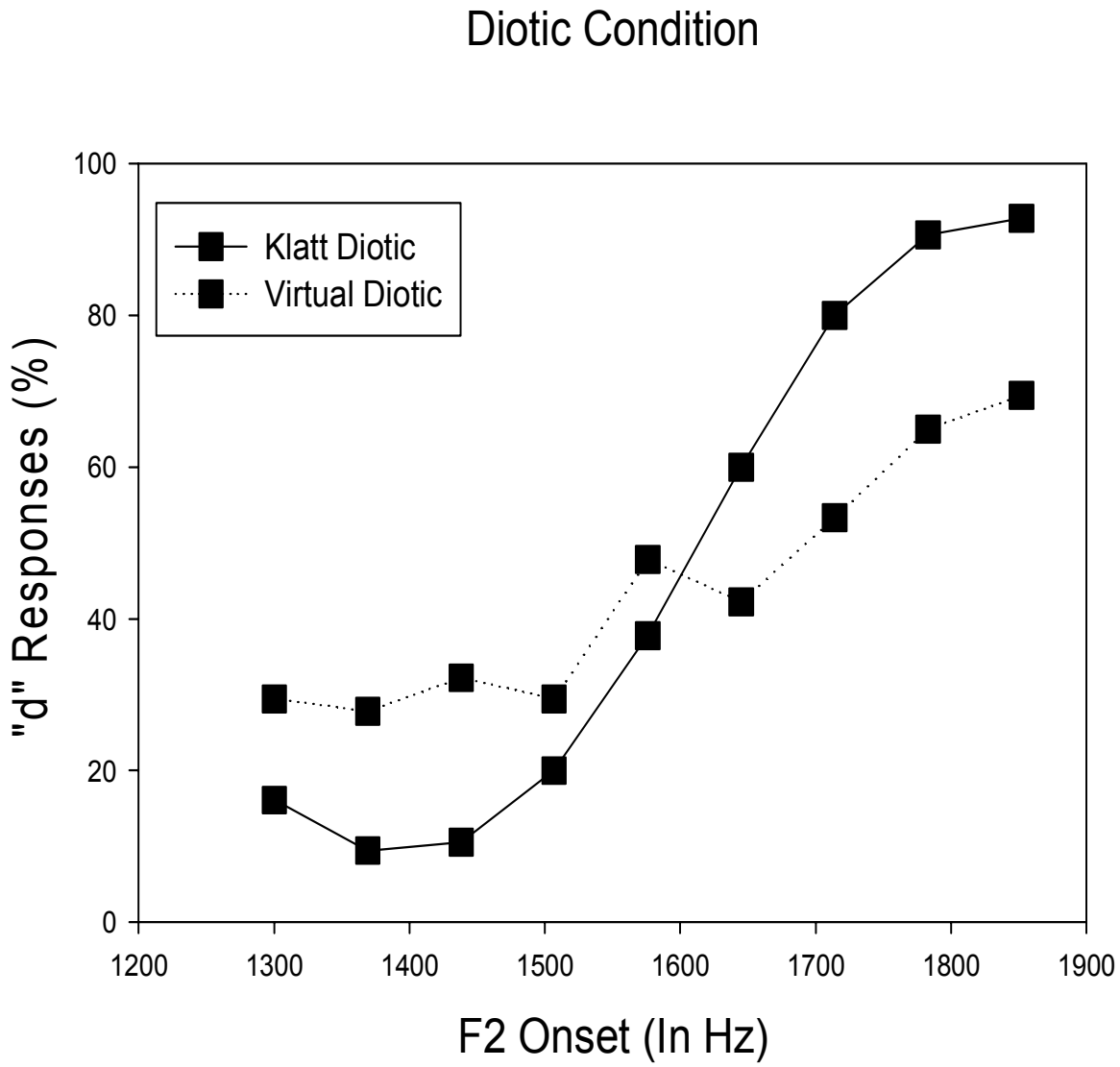

Figure 3.1 The identification functions for the actual and virtual F2 formant transitions in the diotic condition.

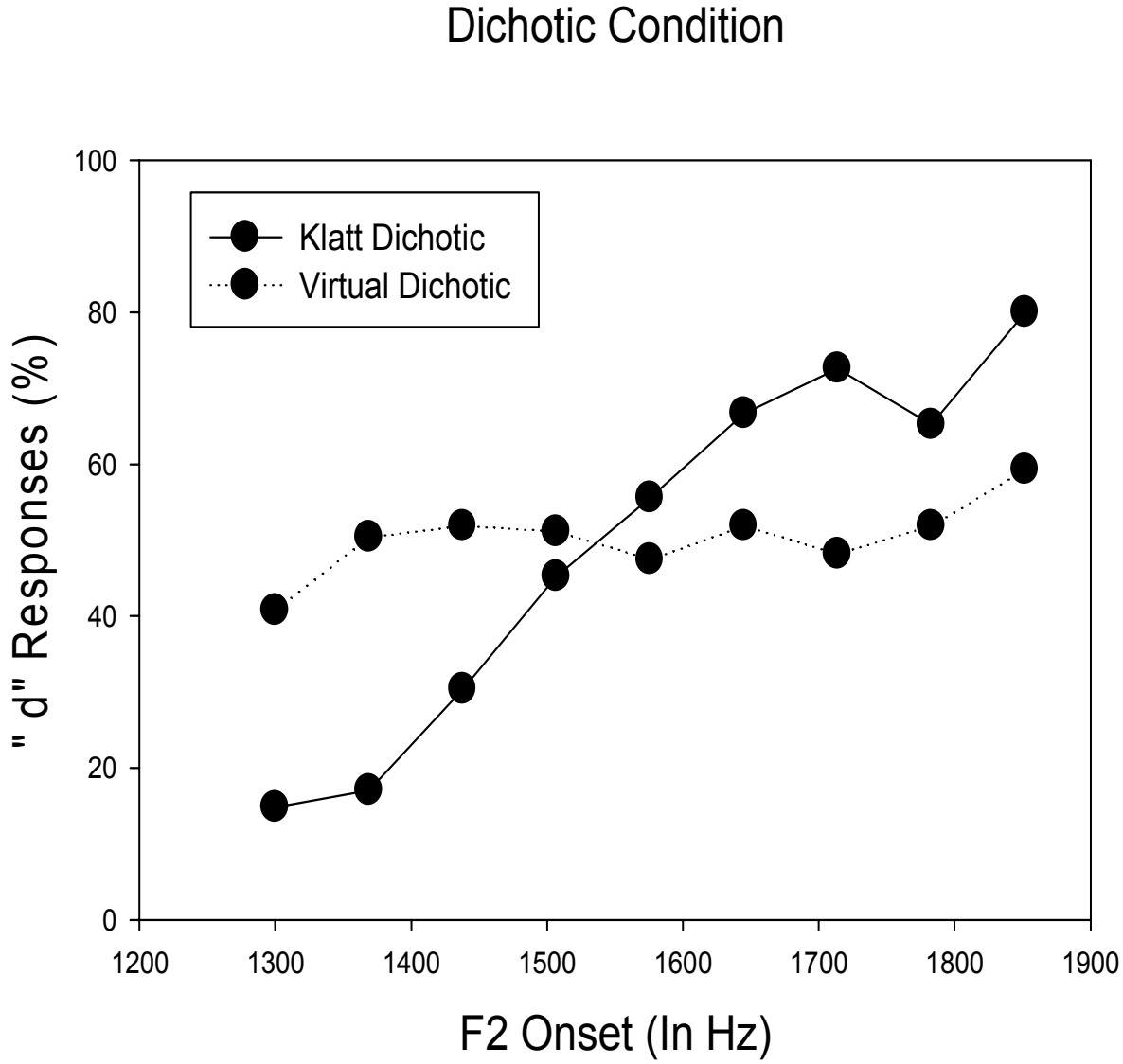

Figure 3.2 The identification functions for the actual and virtual F2 formant transitions in the dichotic condition.

Conclusion

Overall, these results provide evidence that the human auditory system performs auditory spectral integration during speech perception. Datas from both experiments validate the effect of COG in cueing virtual consonant-vowel transition over time. However, the study suggests that actual formants are more salient in cueing /b/-/d/ distinctions regarding place of articulation and that central processing mechanisms might be invoked in the perception of speech and speech-like sounds. Although the dichotic virtual transitions produced a much shallower identification function, there is some slight indication that central processing did take place.

Future studies should further examine the level of spectral integration in auditory processing to determine if it occurs at the peripheral, central or at both levels. An additional issue that also needs to be examined are the intensity value of the virtual F2 transitions. Balancing the perceptual loudness of portions of the stimuli may produce a stronger identification function. The frequency regions of the vowel [æ] should also be analyzed in order to determine the significance of F1 formant and F2 formant location in virtual COG effects. Finally, the absence of an F3 formant should also be addressed. The contributions of F3 information may add spectral richness to the stimuli that might be beneficial for identification. All of these factors need to be examined before one can have a clear picture of spectral integration in auditory processing.

Chapter 4: References

Bedrov, Y.A.; Chistovich, L.A.; and Sheikin, R.L. (1978). Frequency location at the 'center of gravity' of formants as a useful feature in vowel perception. Akust. Zh. 24, 480-486 (Sov. Phys.Acoust., 24, 275-282).

Chistovich, L.A. and Lublinskaja, V.V. (1979). The 'center of gravity' effect in vowel spectra and critical distance between the formants: Psychoacoustical study of the perception of vowel-like stimuli. *Hearing Research,* 1, 185-195.

Chistovich, L.A.; Sheikin, R.L.; Lublinskaja, V.V. (1979). Centres of gravity' and spectral peaks as the determinants of vowel quality, in *Frontiers of Speech Communication Research,* edited by B. Lindblom ans S. Öhman (Academic Press, London), 55-82.

Delattre, P.; Liberman, A.M.; Cooper, F.S.; Gerstman, L.J. (1952). An experimental study of the acoustic determinants of vowel color: Observations on one- and two-formant vowels synthesized from spectrographic patterns. *Word,* 8, 195-210.

Pickett, J.M. (1999). The Acoustics of Speech and Communication, Fundamentals, Speech perception Theory and Technology. Needham Heights, MA . 35-66 and 132-136.

Feth, L.L. (1974). Frequency discrimination of complex periodic tones. *Percept. Psychophys*., 15, 375-378.

Feth, L.L. and O'Malley H. (1977). Two-tone auditory spectral resolution. *J. Acoust. Soc. Am.,* 62, 940-947.

Fox, R.A., Gokcen, J., and Wagner, S. (1997). Neurophysiological and behavioral evidence for a phonetic processor, in *Proceedings from the Panels of the Chicago Linguistic Society's Thirty-third Meeting,* Vol. 33-2, (CLS, Chicago), 311-322.

Fox, R.A., Smith, M., and Jacewicz, E. (2006). Spectral auditory integration and virtual cues to place-of-articulation in stops. *Journal of the Acoustical Society of America,* 119(5), 32-43.

Ladefoged, P. (2001). A Course in Phonetics, fourth ed. Los Angeles: Heinle & Heinle/Thomson Learning.

Lublinskaja, V.V. (1996). The 'center of gravity' effect in dynamics, in *Proceedings of the Workshop on the Auditory Basis of Speech Perception,* W. Ainsworth and S. Greenberg, eds., ESCA, 102-105.

Mann, V.A. and Liberman, A.M. (1983). Some differences between phonetic and auditory modes of perception. *Cognition,* 14, 211-235.

Scharf, B.L. (1972). Critical Bands, in Tobias, J.V., *Foundations of Modern Auditory Theory*, Vol. 1 (Academic Press, New York).

Voelcker, H.B. (1966a). Toward a unified theory of modulation I. Phase-envelope relationship. *Proc. IEEE,* 54, 340-353.

Voelcker, H.B. (1966b). Toward a unified theory of modulation II. Zero manipulation. *Proc. IEEE,* 54, 735-755.

Whalen, D.H. and Liberman, A.M. (1987). Speech perception takes precedence over nonspeech perception. *Science,* 237, 169-171.

Xu, Q.; Jacewicz, E.; Feth, L.L.; Krishnamurthy A.K. (2004). Bandwidth of spectral resolution for two-formant synthetic vowels and two-tone complex signals. *J. Acoust. Soc. Am.,* 115, 1653-1664.

List of Tables and Figures